

Computer Vision, Graphics, and Pattern Recognition Group
Department of Mathematics and Computer Science
University of Mannheim
D-68131 Mannheim, Germany

Reihe Informatik
10/2003

**Natural Image Statistics for Natural Image
Segmentation**

Matthias Heiler, Christoph Schnörr

Technical Report 10/2003
Computer Science Series
July 2003

The publications of the CVGPR Group are listed under
<http://www.cvgpr.uni-mannheim.de>

Natural Image Statistics for Natural Image Segmentation*

Matthias Heiler, Christoph Schnörr

Computer Vision, Graphics, and Pattern Recognition Group
Department of Mathematics and Computer Science
University of Mannheim
D-68131 Mannheim, Germany

Abstract

Building on recent progress in modeling filter response statistics of natural images we integrate a statistical model into a variational framework for image segmentation. Incorporated in a sound probabilistic distance measure the model drives level sets toward meaningful segmentations of complex textures and natural scenes. Despite its enhanced descriptive power our approach preserves the efficiency of level set based segmentation since each connected region comprises two model parameters only. We validate the statistical basis of our model on thousands of natural images and demonstrate that our approach outperforms recent variational segmentation methods based on second-order statistics.

1 Introduction

1.1 Motivation

Statistical models play an increasingly decisive role in computer vision for shape modeling, segmentation, tracking and appearance-based recognition [7]. In the context of Bayesian inference, the nature of a statistical model defines the class of the optimization problem to be solved [16]. As a consequence, there is a trade-off between the descriptive power of statistical models and the difficulty of the associated Bayesian (variational) inference step from the optimization point of view.

Recently, the statistics of filter outputs turned out to provide powerful and general models for image statistics and texture [33, 34, 21, 20, 31]. Unfortunately, incorporating such a model into a variational approach to image segmentation results in a computationally intractable optimization problem which requires the application of time-consuming stochastic sampling methods to compute a minimizer [33, 34, 21].

*A short version of this paper was accepted for ICCV 03.

From the viewpoint of optimization, on the other hand, a range of variational approaches to image segmentation exist [4, 10] which can be regarded as efficient and computationally convenient approximations of the sophisticated Mumford-Shah functional [17]. However, the class of problems to which these models have been applied so far is limited to cartoon-like piecewise smooth images and second-order statistics of multiple filter channel responses [32, 10].

This motivates the use of more sophisticated statistical models which describe a larger class of natural images without compromising the efficiency of level set based segmentation.

1.2 Contribution and Related Work

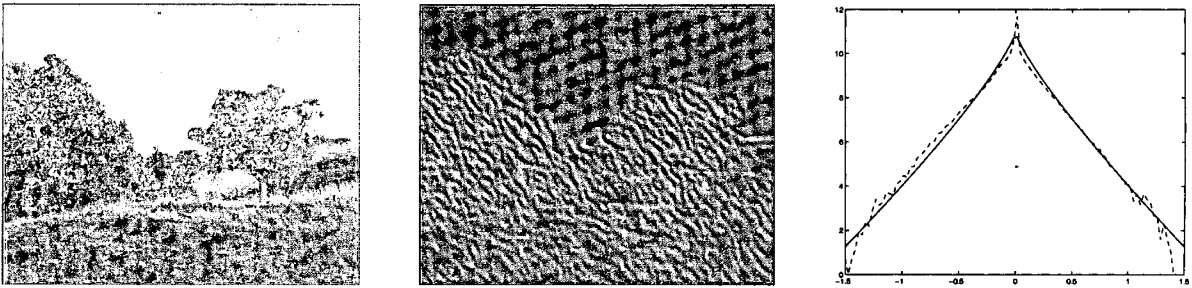
In the present paper, we study for the first time the use of a recent model of natural image statistics in an efficient level set based variational framework for image segmentation.

The statistical model which we describe in Section 2, has been used for wavelet-based image coding [14, 2] and was empirically verified for a large database of natural images [23, 9]. We also refer to [29] for related work.

The variational approach we use for image segmentation follows Zhu and Yuille [32], and Chan and Vese [4]. We considerably enhance the descriptive power of these methods by incorporating the above-mentioned statistical model of natural images (Section 2), thereby enlarging the range of applicability of variational segmentation to a much broader image class. In this sense, our work is similar to Paragios and Deriche [18] who successfully enhance the geodesic active contour model [3, 12] with a Gaussian mixture model of filter response statistics for *supervised* texture segmentation. However, our statistical model is more compact and targeted toward natural scenes, which makes it possible to treat natural images and unknown textures in a completely unsupervised way. We give a rigorous derivation of the equations driving the motion of region boundaries toward a segmentation in Section 3.

In Section 4 we evaluate our approach numerically using natural images and texture images from publicly available databases. The performance is compared to the use of second order statistics as in [32, 10] within the *same* variational framework, both in a supervised and unsupervised setting. It turns out that, despite its enhanced descriptive power, the statistical model we exploit is still simple enough to be estimated *locally* such that, by measuring an appropriate distance to non-locally estimated models, forces can be computed which drive level sets to meaningful segmentations. We conclude and indicate further work in Section 5.

Finally, we wish to point out that our results should not solely be judged from the specific viewpoint of texture segmentation. This would amount to consider a wide range of possible dissimilarity measures [22] many of which cannot be easily incorporated into a levelset based segmentation framework and are not appropriate for less textured parts of natural scenes. Rather, we focus in this paper on a compact parametric model related to natural image statistics for super- and unsupervised levelset segmentation of scenes where texture plays an important but not an exclusive role.



(a) Normalized image

(b) Filter response.

(c) Histogram and model.

Figure 1: **Overview.** An image (a) is filtered by a linear filter. The marginal histogram (dotted line) of the resulting filter response image (b) is computed and a generalized Laplacian (solid line) is fitted to the histogram (c). The parameters (α, s) of the generalized Laplacian serve as image descriptors.

2 Natural Image Statistics

We capture the statistics of natural images using generalized Laplacians fitted to marginal histograms of linear filter responses. The Kullback-Leibler (KL) distance between the Laplacians then serves as a distance measure on the images (cf. Fig. 1). The following section describes the statistical model in detail.

2.1 Feature Extraction

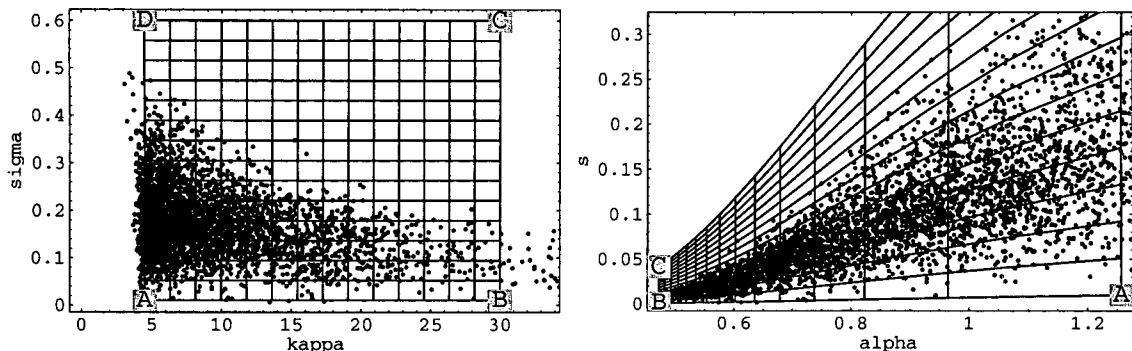
The basis of our approach is the statistical model

$$p(z) = \frac{\alpha}{2s\Gamma(1/\alpha)} \exp(-|z/s|^\alpha) \quad (1)$$

for the filter response z of a linear filter applied to natural images. It was pointed out in the literature that for a large class of images the generalized Laplacian (1) describes the response statistics of various linear filters surprisingly well. This model was empirically verified for samples of natural images [23, 9] and has been applied to compactly code wavelet coefficients of images [14, 11, 2, 13] and to Bayesian image restoration [26]. In the present paper we apply this model both *locally and globally* within a variational framework to the segmentation of natural images.

Various linear transformations of images have been used in conjunction with the model: The discrete cosine transform [23], steerable pyramids [6, 27, 26], and various orthogonal wavelets [14, 9]. In this work we examine steerable pyramid filters and quadrature mirror filters as well as the well-known Haar wavelet and Daubechies wavelet of order 3. In the following these filters are abbreviated by

- spn,
- qmfn,
- haar, and



(a) Filter response parameters.

(b) Density parameters.

Figure 2: The role of nonlinear parameter mapping. Standard deviation σ and kurtosis κ of the filter responses are mapped nonlinearly according to (2) to density parameters α and s . Note that the left part of (a), where most points are located, is spread after mapping. Conversely, the area on the right of (a), where relatively few points are situated, is compressed. The points depicted are 4167 measurements collected from the van Hateren database using a linear derivative filter. The labeled grid visualizes the nonlinearity of the transformation.

- daub3,

where n is an integer encoding the number of orientations examined. Experimental results to be discussed below reveal that both the choice of the filter bank and the metric affect the performance of the statistical model.

2.2 Density Estimation

The generalized Laplacian model (1) has two parameters, s and α , which are related to variance σ^2 and kurtosis κ of the filter response by

$$\sigma^2 = \frac{s^2 \Gamma(3/\alpha)}{\Gamma(1/\alpha)} \quad \kappa = \frac{\Gamma(1/\alpha) \Gamma(5/\alpha)}{\Gamma^2(3/\alpha)}. \quad (2)$$

Figure 2 illustrates the nonlinear mapping of the statistical parameters $\{\sigma, \kappa\}$ to the model parameters $\{s, \alpha\}$ in (1). When $\kappa > 9/5$ we can solve the right equation numerically for α and determine s via the left equation. Mathematically, we cannot model distributions with $\kappa \leq 9/5$ as for $\alpha \rightarrow \infty$ the generalized Laplacian approaches the uniform distribution centered at 0, the kurtosis of which equals $9/5$. This is not a severe restriction, however: In the experimental section (Sec. 4 and Fig. 3) we give strong evidence that such statistics are very rare in natural images.

2.3 MDL-Criterion for Segmentation

Our goal is to partition the image domain Ω into two, maybe multiply-connected, regions Ω_{in} and Ω_{out} separated by a contour \mathcal{C} such that the local image statistics are “close“

to the global statistics within Ω_{in} or Ω_{out} , respectively. More precisely, if pdf_x denotes the statistics of a small window W_x centered at image location x , and if pdf_{in} and pdf_{out} denote the statistics of the interior and exterior regions Ω_{in} and Ω_{out} , respectively, then we want to minimize

$$E_{\text{mdl}}(\Omega_{\text{in}}, \Omega_{\text{out}}) = \int_{\mathcal{C}} ds + \int_{\Omega_{\text{in}}} D(pdf_x || pdf_{\text{in}}) dx + \int_{\Omega_{\text{out}}} D(pdf_x || pdf_{\text{out}}) dx \quad (3)$$

where $D(p||q) = - \int p(z) \log(p(z)/q(z)) dz$ is the Kullback-Leibler (KL) distance between densities p and q .

In order to interpret criterion (3), let us assume for a moment that pdf_{in} , pdf_{out} , and pdf_x model the true densities of their domains perfectly. Minimizing (3) then amounts to minimizing the description length of an image code: A minimal code for W_x has average length $H(pdf_x)$, H denoting Shannon entropy [5]. Encoding W_x using the model for one of the regions Ω_{in} or Ω_{out} requires a code of average length $H(pdf_x) + D(pdf_x || pdf_{\text{in/out}})$. In order to minimize (3), we should assign x to the region with minimal KL-distance to pdf_x . The first integral in (3) measures the length of the separating contour \mathcal{C} ensuring that the membership relation, that is whether a specific point x belongs to Ω_{in} or to Ω_{out} , will be inexpensive to encode.

Note that the KL-distance of two generalized Laplacians p and q with parameters $\{s_p, \alpha_p\}$ and $\{s_q, \alpha_q\}$ can be computed by evaluating (2) for sample estimates on the left hand sides, and by inserting the resulting values for the parameters $\{s_p, \alpha_p, s_q, \alpha_q\}$ into the following expression:

$$D(p||q) = \frac{\left(\frac{s_p}{s_q}\right)^{\alpha_q} \Gamma\left(\frac{1+\alpha_q}{\alpha_p}\right)}{\Gamma\left(\frac{1}{\alpha_p}\right)} + \log \left(\frac{s_q \Gamma\left(1 + \frac{1}{\alpha_q}\right)}{s_p \Gamma\left(1 + \frac{1}{\alpha_p}\right)} \right) - \frac{1}{\alpha_p}. \quad (4)$$

2.4 Combining Filter Responses

Given the statistics for a set of filter responses the question arises how to combine information gathered at different scales and orientations. In this work, we strive for a *generic* measure not optimized for any particular set of textures or filters.

We propose, as a first approximation, to treat the statistics of individual filter responses as statistically independent. Under this assumption the individual KL-distances simply add up so that we can minimize the average distance collected over all linear filters i :

$$E_{\text{mdl}}(\Omega_{\text{in}}, \Omega_{\text{out}}) = \int_{\mathcal{C}} ds + \sum_i \left[\int_{\Omega_{\text{in}}} D(pdf_{x,i} || pdf_{\text{in},i}) dx + \int_{\Omega_{\text{out}}} D(pdf_{x,i} || pdf_{\text{out},i}) dx \right] \quad (5)$$

Here $pdf_{\text{in/out},i}$ denotes the probability density function modeling the response of filter i in region $\Omega_{\text{in/out}}$ and $pdf_{x,i}$ is the corresponding density for a window W_x centered at location x in the image plane.

It is known that in reality the independence assumption does *not* hold. For orthogonal wavelet bases normalization schemes have been proposed to remove dependencies between filter responses at different scale [25]. For the time being, however, we did not

incorporate any normalization schemes into the first implementation of our approach. While this clearly is suboptimal in theory, empirical evidence (Section 4) suggests that our model is sufficiently accurate for a range of real-world scenes.

3 Levelset Formulation

In this section we incorporate our statistical distance measure on images into a levelset formulation. The update equations determining the dynamics of the segmentation are rigorously derived by taking into account all region-dependent terms and by computing the first variation of the corresponding area integrals.

3.1 Energy Functional

Motivated by the region-based segmentation model of Chan and Vese [4] we generalize (3) and consider energy functionals of the form

$$E(\phi) = \int_{\Omega} k^b(x) |\nabla \phi| \delta(\phi) dx + \lambda_1 \int_{\Omega} k^{\text{out}}(x, \phi) H(\phi) dx + \lambda_2 \int_{\Omega} k^{\text{in}}(x, \phi) (1 - H(\phi)) dx \quad (6)$$

Here $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ denotes the embedding level set function, the zero-level of which represents segmentation boundaries. $H : \mathbb{R} \rightarrow \{0, 1\}$ is the step function, and δ stands for Dirac's delta function. In the following we drop the function arguments ϕ and x for brevity.

Chan and Vese's model [4] fits into this framework as a special case

$$\begin{cases} k^b &= 1 \\ k^{\text{in}} &= |u_0 - c_{\text{in}}|^2 \\ k^{\text{out}} &= |u_0 - c_{\text{out}}|^2, \end{cases} \quad (7)$$

whereas our approach involves the considerably more general statistical model from (5)

$$\begin{cases} k^b &= 1 \\ k^{\text{in}} &= \sum_i D(\text{pdf}_{x,i} || \text{pdf}_{\text{in},i}) \\ k^{\text{out}} &= \sum_i D(\text{pdf}_{x,i} || \text{pdf}_{\text{out},i}). \end{cases} \quad (8)$$

3.2 First Variation and Boundary Update

The variational update $\dot{\phi} = -\langle E'(\phi), \psi \rangle, \forall \psi$, of the level set function reads¹:

$$\begin{aligned} \frac{\partial E}{\partial \phi} = \frac{\partial}{\partial \phi} \left[\int_{\Omega} k^b |\nabla \phi| \delta dx \right] + \int_{\Omega} (\lambda_1 k^{\text{out}} - \lambda_2 k^{\text{in}}) \delta \psi dx \\ + \int_{\Omega} \left(\lambda_1 \frac{\partial k^{\text{out}}}{\partial \phi} H + \lambda_2 \frac{\partial k^{\text{in}}}{\partial \phi} (1 - H) \right) \psi dx. \end{aligned} \quad (9)$$

¹To save horizontal space we abbreviate $\langle E'(\phi), \psi \rangle$ by $\partial E / \partial \phi$.

The third term, which is omitted in [4], originates from applying the product rule to the area integrals and thus takes into account that k^{in} and k^{out} also depend on the levelset function ϕ . After some tedious calculations (Appendix A) and with the shorthands $n = \frac{\nabla\phi}{|\nabla\phi|}$ and $c = \text{div}(n)$ we arrive at

$$\begin{aligned} \frac{\partial E}{\partial \phi} = & \int_c (-\nabla k^{\text{b}} n - k^{\text{b}} c + \lambda_1 k^{\text{out}} - \lambda_2 k^{\text{in}}) \psi \, ds \\ & + \int_{\Omega} \left(\lambda_1 \frac{\partial k^{\text{out}}}{\partial \phi} H + \lambda_2 \frac{\partial k^{\text{in}}}{\partial \phi} (1 - H) \right) \psi \, dx. \end{aligned} \quad (10)$$

We point out that this formula was recently derived in a different way in [10] based on the calculus of shape optimal design [28] which, in turn, relies on previous mathematical work like, e.g., [24].

3.3 Computation of the Area Term

Let us examine more closely the area integral in (10). As mentioned above (Section 2.4) we model the local coding cost w.r.t. the interior region as

$$k^{\text{in}} = \sum_i D(\text{pdf}_{x,i} || \text{pdf}_{\text{in},i}). \quad (11)$$

Recall that the probability density functions is given as generalized Laplacians with two parameters $s = s(\alpha, \sigma^2)$ and $\alpha = \alpha(\kappa)$ which depend themselves on kurtosis κ and variance σ^2 measured both locally in W_x and globally in Ω_{in} . Therefore, we may write more precisely

$$k^{\text{in}} = \sum_i D(p(\alpha(\kappa_{x,i}), s(\alpha(\kappa_{x,i}), \sigma_{x,i}^2)) || p(\alpha(\kappa_{\text{in},i}), s(\alpha(\kappa_{\text{in},i}), \sigma_{\text{in},i}^2))). \quad (12)$$

Here $\kappa_{\text{in},i}$ and $\sigma_{\text{in},i}^2$ depend on the area Ω_{in} and thus vary with the levelset function ϕ . Let us drop the index i in the following discussion, thus focusing on a single filter response only.

The derivative then reads

$$\frac{\partial k^{\text{in}}}{\partial \phi} = \frac{\partial D}{\partial \kappa_{\text{in}}} \frac{\partial \kappa_{\text{in}}}{\partial \phi} + \frac{\partial D}{\partial \sigma_{\text{in}}^2} \frac{\partial \sigma_{\text{in}}^2}{\partial \phi}, \quad (13)$$

where the computation of the partial derivatives $\partial D / \partial \kappa_{\text{in}}$ and $\partial D / \partial \sigma_{\text{in}}^2$ is long but nevertheless elementary when starting from the analytical formulation of the KL-distance (4) and inserting the relations (2) solved for α and s .

The statistics depending on the area form a hierarchy of region-dependent terms:

$$\begin{aligned} \kappa_{\text{in}} &= \int_{\Omega_{\text{in}}} \frac{(x - \mu_{\text{in}})^4}{|\Omega_{\text{in}}| \sigma_{\text{in}}^4} dx & \sigma_{\text{in}}^2 &= \int_{\Omega_{\text{in}}} \frac{(x - \mu_{\text{in}})^2}{|\Omega_{\text{in}}|} dx \\ \mu_{\text{in}} &= \int_{\Omega_{\text{in}}} \frac{x}{|\Omega_{\text{in}}|} dx & |\Omega_{\text{in}}| &= \int_{\Omega_{\text{in}}} dx. \end{aligned} \quad (14)$$

In the levelset formulation (6) we replace the integrals over Ω_{in} by integrals over Ω weighted by the step function H . Now, taking the derivative w.r.t. ϕ yields (cf. Appendix B)

$$\frac{\partial \sigma_{\text{in}}^2}{\partial \phi} = - \int_{\Omega} \frac{(x - \mu_{\text{in}})^2}{|\Omega_{\text{in}}|^2} H \, dx \int_{\Omega} \delta \psi \, dx + \int_{\Omega} \frac{(x - \mu_{\text{in}})^2}{|\Omega_{\text{in}}|} \delta \psi \, dx \quad (15)$$

and

$$\begin{aligned} \frac{\partial \kappa_{\text{in}}}{\partial \phi} = & \int_{\Omega} \frac{-4(x - \mu_{\text{in}})^3}{|\Omega_{\text{in}}| \sigma_{\text{in}}^4} H \, dx \left[\int_{\Omega} \frac{-x}{|\Omega_{\text{in}}|^2} H \, dx \int_{\Omega} \delta \psi \, dx + \int_{\Omega} \frac{x}{|\Omega_{\text{in}}|} \delta \psi \, dx \right] \\ & + 2\sigma_{\text{in}}^2 \int_{\Omega} \frac{(x - \mu_{\text{in}})^4}{|\Omega_{\text{in}}| \sigma_{\text{in}}^8} H \, dx \left[\int_{\Omega} \frac{(x - \mu_{\text{in}})^2}{|\Omega_{\text{in}}|^2} H \, dx \int_{\Omega} \delta \psi \, dx - \int_{\Omega} \frac{(x - \mu_{\text{in}})^2}{|\Omega_{\text{in}}|} \delta \psi \, dx \right] \\ & - \int_{\Omega} \frac{(x - \mu_{\text{in}})^4}{|\Omega_{\text{in}}|^2 \sigma_{\text{in}}^4} H \, dx \int_{\Omega} \delta \psi \, dx + \int_{\Omega} \frac{(x - \mu_{\text{in}})^4}{|\Omega_{\text{in}}| \sigma_{\text{in}}^4} \delta \psi \, dx. \end{aligned} \quad (16)$$

With (13) these terms form the area derivatives in (10).

4 Experiments and Discussion

In this we describe extensive computational studies of the performance of our model. We validate the use of generalized Laplacian densities for filter response statistics of natural images, perform experiments in texture retrieval and -synthesis to understand what image features are captured by our model, and show sample segmentations on natural and artificial images. We compare our model with a standard second-order variational model for image segmentation and demonstrate that it performs well.

4.1 Model Validation

Before focusing on segmentation (Section 4.3) we conducted experiments to select a suitable filter bank and to verify that the mathematical restriction on the kurtosis of the filter response to be greater than 9/5 is met in practice (cf. Section 2.2). Following [9] we used the van Hateren database of natural images [30] for evaluation and removed multiplicative constants from the images by first log-transforming them and then subtracting their log-means.

Table 1 summarizes our results: We display the median and the two quartiles of the KL-distance between the filter response histograms and a generalized Laplacian with identical variance and kurtosis. For comparison, we also report the histograms' average entropy. *The results show that on average almost all information in the histograms is captured by the parametric model.*

The following experiments were conducted using the steerable pyramid filter bank sp3 with four oriented sub-band filters and three scales. In Fig. 3 we show the log-histograms of the kurtosis κ for each individual filter determined for all 4167 images of the database. Two things are remarkable: First, the distribution of κ follows closely a

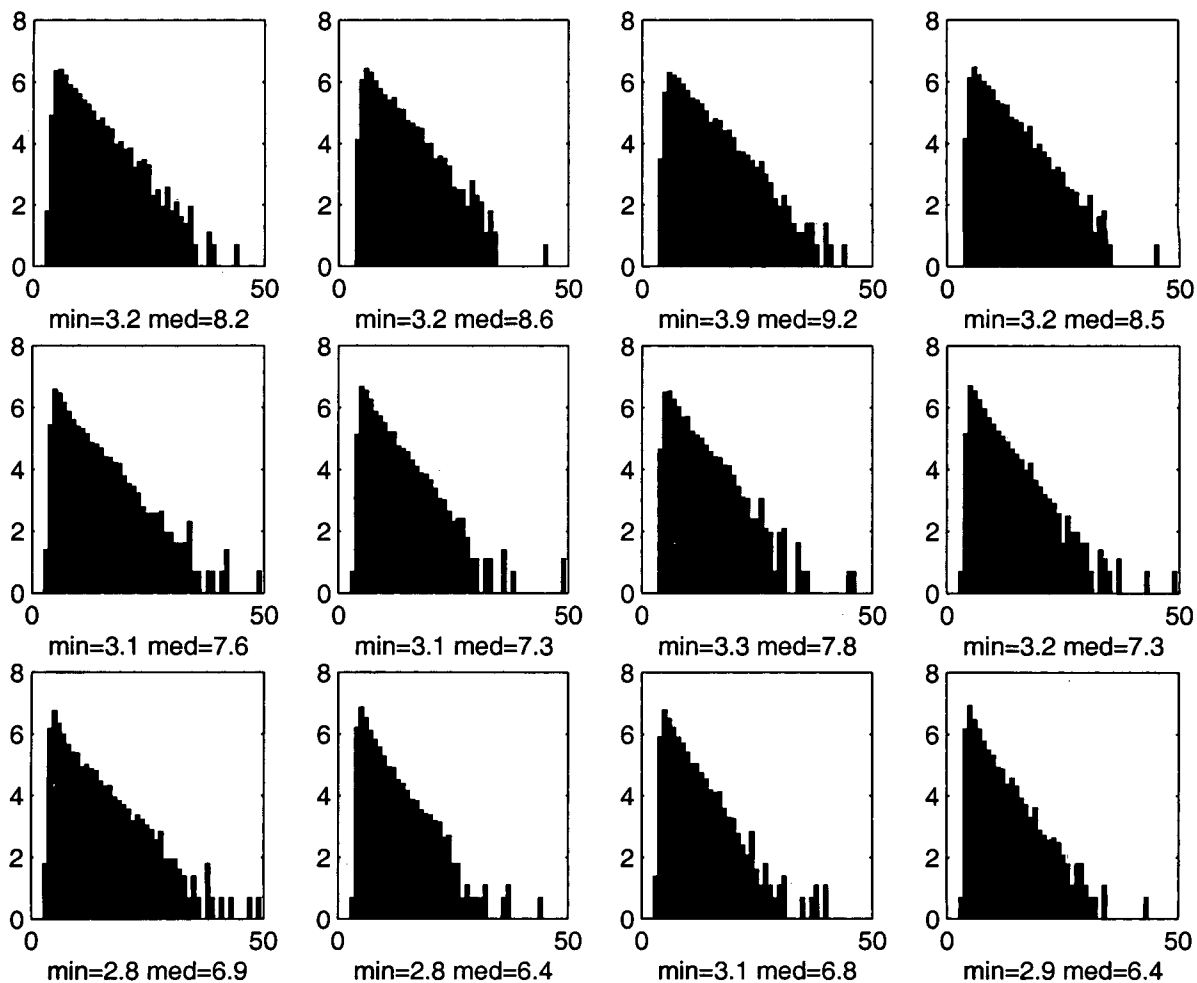


Figure 3: **Check for pathological statistics.** Log-histogram of kurtosis κ measured over 4167 images from the van Hateren database [30] for a steerable pyramid filter bank with three scales (rows) and four orientations (columns). The minimal and median values for κ are listed in the individual image captions. Note that the histograms are *very regular*, and that for each filter κ is well above $9/5$, thus *no pathological cases* are present in the database.

	sp0	sp1	sp3	sp5	qmf9	qmf12	qmf16	haar	daub3
KL-dist median	0.053	0.056	0.058	0.051	0.126	0.126	0.124	0.103	0.111
KL-dist q-25	0.033	0.037	0.040	0.036	0.074	0.074	0.075	0.064	0.068
KL-dist q-75	0.088	0.091	0.090	0.076	0.264	0.263	0.251	0.191	0.213
entropy median	4.498	4.286	4.220	4.223	3.829	3.824	3.832	3.868	3.857
entropy q-25	4.095	3.873	3.828	3.820	3.458	3.447	3.461	3.473	3.489
entropy q-75	4.801	4.608	4.521	4.527	4.138	4.135	4.142	4.200	4.165
KL-dist/entropy	0.011	0.013	0.013	0.012	0.033	0.033	0.032	0.026	0.028

Table 1: **Model validation.** Median and quartiles of KL-distances between histograms and parametric model (1) measured over 4167 pictures from the van Hateren database [30] for different sets of filters. For comparison, entropies of the filter responses are also reported. The steerable pyramid filter responses (columns 1 to 4) fit particularly well: Only a small fraction of the information present in the histograms is ignored (last row).

shifted exponential distribution. Second, the minimal values of κ encountered are well above the critical value of $9/5$. Thus, *distributions which cannot be described by our model do not occur in natural images.*

Unfortunately, during segmentation we work with small *parts* of images for which small values for kurtosis *are* observed. Especially very homogeneous image regions like sky or plain street occasionally lead to untypical filter response histograms. In the segmentation experiments reported below we treated these histograms as outliers — without any noticeable deterioration of segmentation quality.

To assess the quality of our probabilistic distance measure we ran an experiment in texture retrieval on images from the Brodatz database: We extracted 16 image patches of size 100×100 pixels non-overlappingly from 32 Brodatz images. Then we took each of the $32 \cdot 16 = 512$ image patches as a query and selected from the remaining patches the one most similar to the query w.r.t. a number of norms: We examined L_1 and L_∞ norms for vectors of filter response statistics collected over different scales and orientations and for different sets of linear filters. Also, we examined KL-distance. The distances for the individual scales and orientations were combined using the max, mean, and median operator. A retrieval was considered correct if the patch most similar to the query originated from the same Brodatz image. Otherwise it was considered wrong.

Table 2 summarizes the results which show that the L_1 norm of the KL-distance performs best for most filters. *This indicates that (12) is a useful distance measure on images with texture.*

	L_∞	L_1	med	L_∞	L_1	med	L_∞	L_1	med
	(σ, κ)	(σ, κ)	(σ, κ)	(α, s)	(α, s)	(α, s)	(KL)	(KL)	(KL)
sp0	112	93	120	136	137	303	75	71	103
sp1	64	32	60	66	51	111	33	26	34
sp3	62	30	40	66	48	47	29	18	24
sp5	77	32	33	65	32	45	38	30	24
qmf9	64	19	32	66	38	82	16	12	17
qmf12	79	19	35	78	39	89	20	16	23
qmf16	84	21	34	84	41	83	20	17	22
haar	78	28	39	89	43	122	34	21	35
daub3	82	23	44	80	46	90	29	17	22

Table 2: **Assessment of KL-distance.** Number of false retrievals on 512 randomly sampled texture image patches from the Brodatz database [1]. Different filters, features, and strategies to combine individual filter responses are compared. The minimal error rate for each filter is marked in boldface. KL-distance with L_1 metric performs best in most cases.

4.2 Texture Synthesis Experiment

To understand which image features are captured we synthesized texture images using our model. For computational efficiency we did not resort to the Gibbs sampler but modified the fast pyramid-based algorithm of Heeger and Bergen [8] instead. This greedy algorithm enforces filter histogram similarity between a target image and a source image initialized to random noise over different scale and orientation bands of a steerable pyramid. In contrast to Heeger and Bergen we did not fit the complete filter histograms but only their generalized Laplacians. A similar approach was taken in [29] where Bessel K forms and the Gibbs sampler were used to synthesize texture images from a larger number of linear filter responses.

Figure 4 shows some results: While our – from the viewpoint of image synthesis: simple – method does not produce realistically looking textures it appears that essential information for image segmentation such as scale, granularity and predominant orientation are retained.

4.3 Segmentation Experiments

Next we composed randomly selected textures from the Brodatz database and arranged them in a texture collage with a cross-shaped inlay of one texture in another (Fig. 5). We segmented 100 texture collages using (9) *without area derivatives* and with fixed default parameters $\lambda_1 = \lambda_2 = 1$, window size $|W| = 80 \times 80$ pixels. For comparison, we

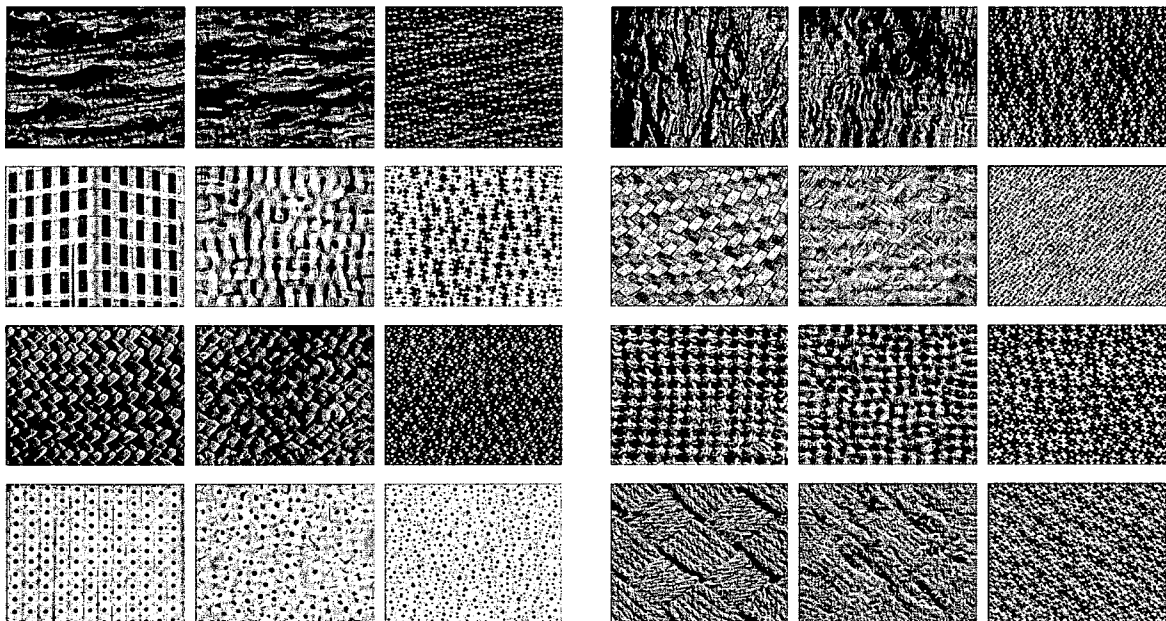


Figure 4: **Texture synthesis.** Textures from the VisTex database (first and fourth column) are reproduced using the algorithm of Heeger and Bergen [8] (second and fifth column) and a simplified version which uses only image features captured by our model (third and sixth column). All images were synthesized using identical filters and the same number of iterations. We see that our model reproduces scale, granularity and predominant orientation of the image.

implemented an energy term based on second order statistics (cf.[32, eqn. (20)]):

$$\begin{cases} k^b = 1 \\ k^{\text{in}} = \sum_i \log(\sigma_{\text{in},i}^2) + \frac{(\mu_{x,i} - \mu_{\text{in},i})^2}{\sigma_{\text{in},i}^2} + \sigma_{x,i}^2 / \sigma_{\text{in},i}^2 \\ k^{\text{out}} = \sum_i \log(\sigma_{\text{out},i}^2) + \frac{(\mu_{x,i} - \mu_{\text{out},i})^2}{\sigma_{\text{out},i}^2} + \sigma_{x,i}^2 / \sigma_{\text{out},i}^2. \end{cases} \quad (17)$$

This model should work well for images where the mean is the most important region descriptor (Fig. 8(h)). Our Brodatz-collages are of such type: The individual texture images usually are quite homogeneous so filter response differences are likely to origin from texture boundaries.

We ran both models for 100 iteration steps, i.e., well after we expected convergence, on each texture collage. For increased speed we computed the image statistics on a subsampled image and interpolated the result on the whole image. This makes the region boundaries look slightly smoother than one would expect. As both models are affected in exactly the same way this should not affect the model comparison. We finally determined the percentage of correctly segmented pixels. We found (Tab. 3) that the average performance (median) as well as the performance on difficult images (25% quartile) of our model was significantly better than the performance of model (17).

We then evaluated the importance of the area derivatives, which are often omitted

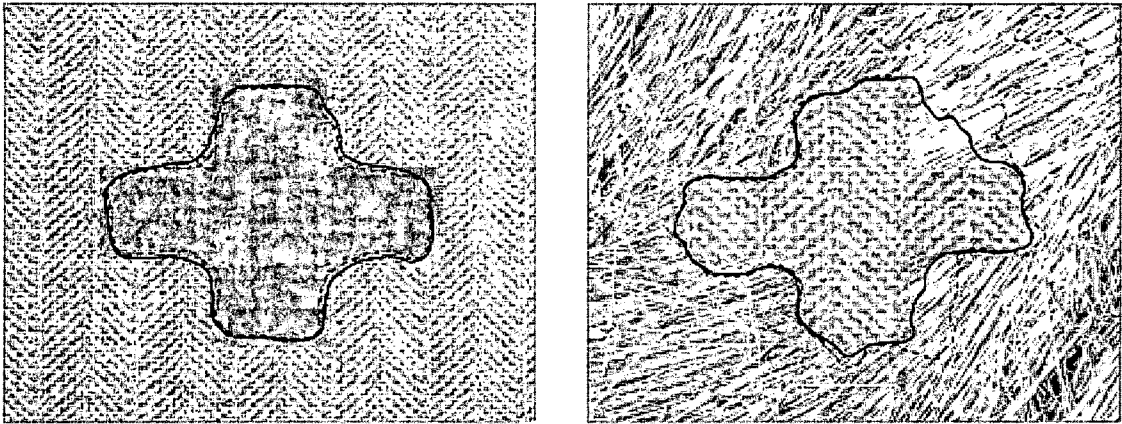


Figure 5: **Sample segmentation.** Brodatz texture collages segmented with KL-distance (solid line) and second order statistics (dotted line) with default parameters set. The KL-distance captures the cross-shaped inlay better than second order statistics. Here, we show two examples out of a large number of segmentation experiments, the statistics of which is given in table 3.

	reference model	proposed model	improvement
median	0.65	0.81	25%
q-25	0.47	0.69	47%
q-75	0.81	0.84	4%

Table 3: **Comparison of segmentation quality.** The percentage of correctly segmented pixels on a set of 100 randomly generated Brodatz texture collages is reported for our model and for a reference model based on second order statistics. The median and both quartiles are shown. Our model clearly outperforms the reference model on average and shows much better performance on difficult images.

in variational segmentation implementations. We took the first 100 images from the van Hateren database and computed the area derivative term from (9)

$$\lambda_1 \frac{\partial k^{\text{out}}}{\partial \phi} H + \lambda_2 \frac{\partial k^{\text{in}}}{\partial \phi} (1 - H) \quad (18)$$

for an initial segmentation consisting of equally spaced squares distributed over the whole image (Fig. 8(a)). For comparison, we computed the KL-term

$$\lambda_1 k^{\text{out}} - \lambda_2 k^{\text{in}}, \quad (19)$$

and measured the influence over the whole image.

The results (Tab. 4) indicate clearly that for our choice of distance measure *the area derivatives can safely be ignored*. This validates common practice. Note, however, that

	q-10	q-90
KL-term	-4.0366	2.6027
Area-term	$-3.61 \cdot 10^{-5}$	$2.11 \cdot 10^{-5}$

Table 4: **Importance of the area term.** The 10% and the 90% quantiles of equations (18) and (19) evaluated on 100 images from the van Hateren database are reported. The contributions of the area term are five orders of magnitude smaller than the contributions of the KL-term, indicating that the area derivatives can safely be ignored.

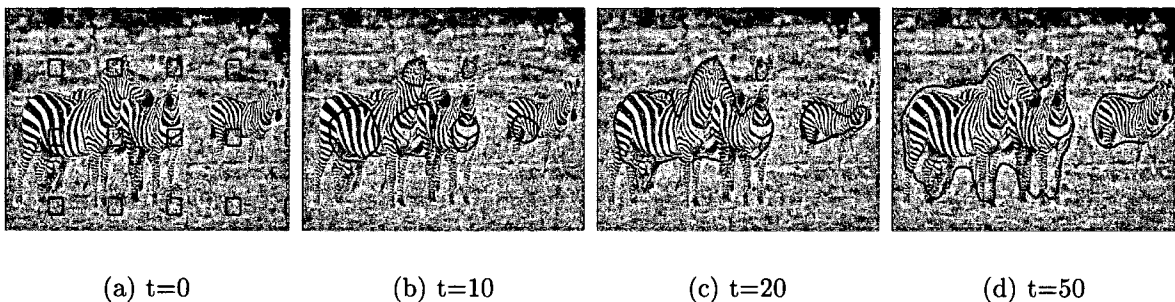


Figure 6: **Unsupervised segmentation.** Zebras are separated from the background. Contours were initialized to boxes, stopping was determined automatically according to $E(\phi)'$.

recently Jehan-Besson et al. [10] have come to different conclusions for a different choice of distance measure.

Fig. 6 to 8 show some examples for supervised and unsupervised segmentation of natural images. In Fig. 6 we examine an image from the Berkeley database [15]. The contour was initialized to equally spaced boxes. As stopping criterion we computed the improvement of the energy functional (6) for every time step and stopped as soon as it dropped below a previously determined threshold. The same threshold was used for all experiments. The zebra pattern is captured well by our model: The contour immediately locks onto the zebra pattern and energy (6) (not shown) drops sharply as the zebra is more and more covered.

Fig. 7 shows a more difficult case: A tree standing in front of a house, casting a sharp shadow on the house. With this image, unsupervised segmentation merely separates the irregular regions from the homogeneous sky and parts of the streets (Fig. 7(e) – Fig. 7(h)). In contrast, if the contour is initialized in a supervised way (Fig. 7(a)) the model captures the visually dominant tree. However, in the final segmentation (Fig. 7(d)) relatively large parts of the shadowed house are captured as well.

In Fig. 8 we compare our model to second order statistics (18) on an image from the MIT VisTex database [19]. The MDL criterion (3) separates the trees from the image fore- and background. This is sensible: The trees form an image region which is relatively expensive to encode while sky and grassland are comparatively homogeneous. Using one probability model for the trees and one for the rest of the image thus minimizes

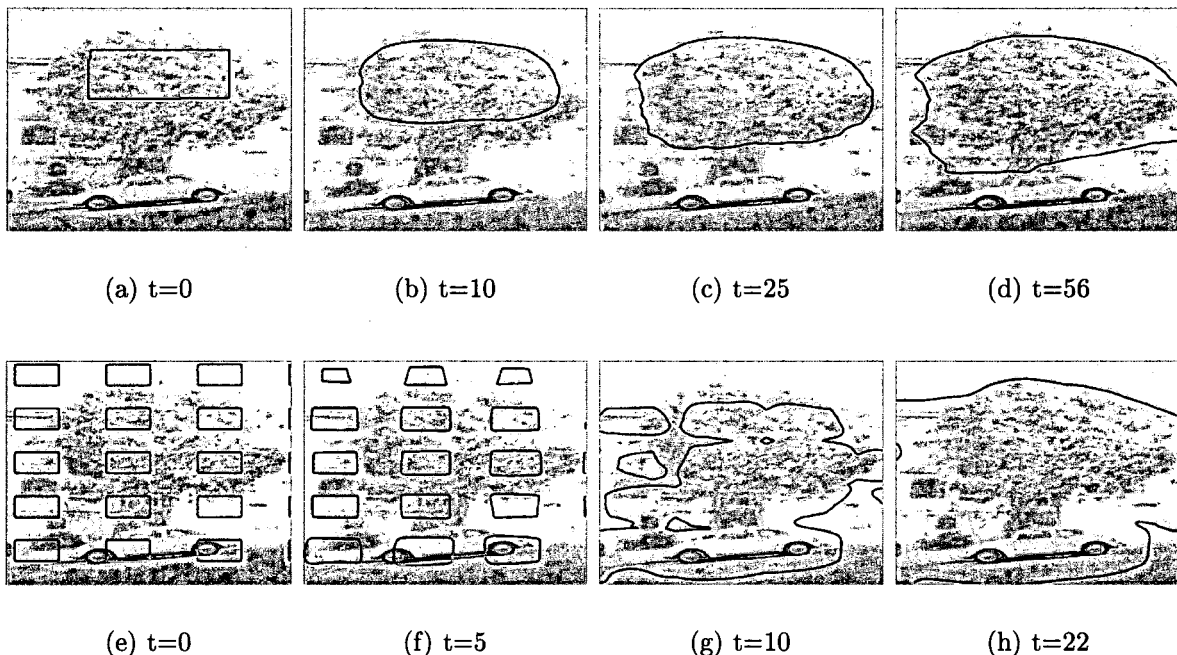


Figure 7: **Supervised and unsupervised segmentation.** With supervised segmentation the tree is separated from house and car. Unsupervised segmentation fails in this case: Initialization of the filter response model is too unspecific, yielding a rather uninteresting segmentation into homogeneous (sky, street) and inhomogeneous regions (car, tree, house). Note the low image contrast in the lower left part of the tree.

the expected coding length of the image. Second order statistics simply separates the bright sky from the rest of the image, yielding a less appealing segmentation.

5 Conclusions and Further Work

In this paper we proposed a segmentation approach based on natural image statistics and the gradient-less level set segmentation method introduced by Chan and Vese [4]. Exploiting the fact that a simple parametric model accurately describes the statistics of a wide class of filter responses on natural images we constructed an energy functional justified by a minimum description length argument.

We ran evaluations on thousands of images checking that pathological cases not captured by our model do not occur in real world images (Fig. 3), that the empirically observed histograms are accurately represented (Tab. 1), and that the minimum description length formulation does contribute to the descriptive power of our model (Tab. 2).

We conducted experiments to evaluate the performance of our segmentation method in comparison to a second order model which has been used successfully for image segmentation before [32]. The results indicate that for segmentation tasks where image structure is more important than brightness contrasts our model compares favorably (Fig. 7, 8, 5; Tab. 3).

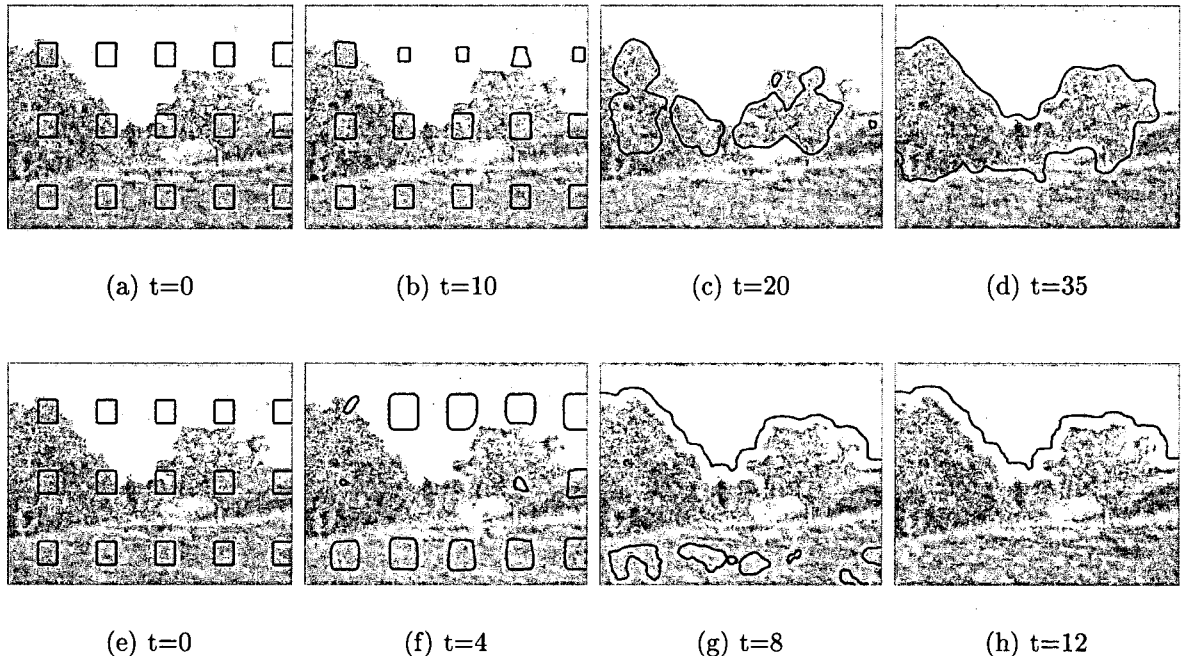


Figure 8: **Unsupervised segmentation.** Unsupervised segmentation of a natural scene from the VisTex database [19]. Contours were initialized to boxes, stopping was determined automatically according to $E(\phi)'$. The contour evolution at different time steps is displayed for our model (Fig. (a)–(d)) and for second order statistics (Fig. (e)–(h)). The trees in the center of the image are the visually most dominant element which is reflected by the segmentation with our model. Second order statistics separates the bright sky from the darker rest of the image, failing to capture the visually dominant trees.

Finally, we examined the importance of an area derivative term emerging during the derivation of the first variation of our energy functional. We found (Tab. 4) that for our functional the area derivative’s contribution is not significant, thus validating the common practice of omitting it. Omitting the area derivatives also greatly simplifies the implementation of our method and lifts the requirement that the area descriptors k^{in} and k^{out} must be differentiable.

An interesting line of research for the future is to examine how image features can be captured with even more involved image probability densities without compromising model simplicity too much: Our energy functional can in principle be applied on arbitrary probability densities. However, model validation issues as well as performance arguments make image models desirable which are easy to train even on small image patches.

Acknowledgments

The authors would like to thank Eero P. Simoncelli for making his steerable pyramid toolbox publicly available.

A First Variation of the Energy Functional

The exposition follows [4], suitably generalized and adapted to our approach. Starting from (9)

$$\begin{aligned} \frac{\partial E}{\partial \phi} = \frac{\partial}{\partial \phi} \left[\int_{\Omega} k^b |\nabla \phi| \delta dx \right] + \int_{\Omega} \delta(\lambda_1 k^{\text{out}} - \lambda_2 k^{\text{in}}) \psi dx \\ + \int_{\Omega} \left(\lambda_1 \frac{\partial k^{\text{out}}}{\partial \phi} H + \lambda_2 \frac{\partial k^{\text{in}}}{\partial \phi} (1 - H) \right) \psi dx \end{aligned}$$

we take a closer look at the first term which equals

$$\int_{\Omega} k^b \left[\delta' |\nabla \phi| \psi + \delta \frac{\nabla \phi}{|\nabla \phi|} \nabla \psi \right] dx \quad (20)$$

by product rule. With Green's first theorem the second part becomes

$$\begin{aligned} \int_{\Omega} k^b \delta \frac{\nabla \phi}{|\nabla \phi|} \nabla \psi dx &= - \int_{\Omega} \nabla \left(k^b \delta \frac{\nabla \phi}{|\nabla \phi|} \right) \psi dx + \int_{\partial \Omega} \frac{k^b \delta \partial \phi}{|\nabla \phi| \partial n} \psi ds \\ &= - \int_{\Omega} \left[\nabla k^b \delta \frac{\nabla \phi}{|\nabla \phi|} + k^b \nabla \delta \frac{\nabla \phi}{|\nabla \phi|} + k^b \delta \nabla \left(\frac{\nabla \phi}{|\nabla \phi|} \right) \right] \psi dx + \int_{\partial \Omega} \frac{k^b \delta \partial \phi}{|\nabla \phi| \partial n} \psi ds \end{aligned} \quad (21)$$

which in connection with $\nabla \delta \frac{\nabla \phi}{|\nabla \phi|} = \delta' |\nabla \phi|$ and (20) yields

$$\frac{\partial}{\partial \phi} \left[\int_{\Omega} k^b |\nabla \phi| \delta dx \right] = - \int_{\Omega} \left[\nabla k^b \delta \frac{\nabla \phi}{|\nabla \phi|} + k^b \delta \nabla \left(\frac{\nabla \phi}{|\nabla \phi|} \right) \right] \psi dx + \int_{\partial \Omega} \frac{k^b \delta \partial \phi}{|\nabla \phi| \partial n} \psi ds. \quad (22)$$

Note that we can replace each area integral containing the Dirac impulse into an integral over the region boundary $\mathcal{C} = \{x : \phi(x) = 0\}$:

$$\int_{\Omega} f(x, \phi) \delta(\phi) dx = \int_{\mathcal{C}} f(x, 0) ds. \quad (23)$$

Hence we can write

$$\begin{aligned} \frac{\partial E}{\partial \phi} &= \int_{\mathcal{C} \cap \partial \Omega} \frac{k^b \partial \phi}{|\nabla \phi| \partial n} \psi ds + \int_{\mathcal{C}} \left[-\nabla k^b \frac{\nabla \phi}{|\nabla \phi|} - k^b \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + \lambda_1 k^{\text{out}} - \lambda_2 k^{\text{in}} \right] \psi ds \\ &\quad + \int_{\Omega} \left(\lambda_1 \frac{\partial k^{\text{out}}}{\partial \phi} H + \lambda_2 \frac{\partial k^{\text{in}}}{\partial \phi} (1 - H) \right) \psi dx. \end{aligned} \quad (24)$$

Assuming $\mathcal{C} \cap \partial \Omega = \emptyset$, this leads to (10).

B Derivation of the Area Terms

We start from relations (14) and replace the integrals over Ω_{in} by integrals over Ω weighted by the step function H . Taking the derivative w.r.t. ϕ yields

$$\frac{\partial \sigma_{\text{in}}^2}{\partial \phi} = \frac{\partial}{\partial \phi} \int_{\Omega} \frac{(x - \mu_{\text{in}})^2}{|\Omega_{\text{in}}|} H \, dx = \int_{\Omega} \frac{\partial}{\partial \phi} \left[\frac{(x - \mu_{\text{in}})^2}{|\Omega_{\text{in}}|} \right] H + \frac{(x - \mu_{\text{in}})^2}{|\Omega_{\text{in}}|} \delta\psi \, dx \quad (25)$$

and

$$\frac{\partial}{\partial \phi} \left[\frac{(x - \mu_{\text{in}})^2}{|\Omega_{\text{in}}|} \right] = 2 \frac{\mu_{\text{in}} - x}{|\Omega_{\text{in}}|} \frac{\partial \mu_{\text{in}}}{\partial \phi} - \frac{(x - \mu_{\text{in}})^2}{|\Omega_{\text{in}}|^2} \frac{\partial |\Omega_{\text{in}}|}{\partial \phi} \quad (26)$$

and finally

$$\frac{\partial |\Omega_{\text{in}}|}{\partial \phi} = \int_{\Omega} \frac{\partial H}{\partial \phi} dx = \int_{\Omega} \delta\psi \, dx. \quad (27)$$

Collecting these terms and using $\int_{\Omega} (\mu_{\text{in}} - x) H \, dx = 0$ yields (15).

The derivation of $\partial \kappa / \partial \phi$ proceeds in the very same manner:

$$\begin{aligned} \frac{\partial \kappa_{\text{in}}}{\partial \phi} &= \frac{\partial}{\partial \phi} \int_{\Omega} \frac{(x - \mu_{\text{in}})^4}{|\Omega_{\text{in}}| \sigma_{\text{in}}^4} H \, dx = \int_{\Omega} \frac{\partial}{\partial \phi} \left[\frac{(x - \mu_{\text{in}})^4}{|\Omega_{\text{in}}| \sigma_{\text{in}}^4} \right] H + \frac{(x - \mu_{\text{in}})^4}{|\Omega_{\text{in}}| \sigma_{\text{in}}^4} \delta\psi \, dx \\ &= \int_{\Omega} \frac{-4(x - \mu_{\text{in}})^3}{|\Omega_{\text{in}}| \sigma_{\text{in}}^4} \frac{\partial \mu_{\text{in}}}{\partial \phi} H - \frac{(x - \mu_{\text{in}})^4}{|\Omega_{\text{in}}|^2 \sigma_{\text{in}}^8} \left(\sigma_{\text{in}}^4 \frac{\partial |\Omega_{\text{in}}|}{\partial \phi} + |\Omega_{\text{in}}| \frac{\partial \sigma_{\text{in}}^4}{\partial \phi} \right) H + \frac{(x - \mu_{\text{in}})^4}{|\Omega_{\text{in}}| \sigma_{\text{in}}^4} \delta\psi \, dx \end{aligned} \quad (28)$$

where

$$\begin{aligned} \frac{\partial \mu_{\text{in}}}{\partial \phi} &= \frac{\partial}{\partial \phi} \int_{\Omega} \frac{x}{|\Omega_{\text{in}}|} H \, dx = \int_{\Omega} \frac{-x}{|\Omega_{\text{in}}|^2} \frac{\partial |\Omega_{\text{in}}|}{\partial \phi} H + \frac{x}{|\Omega_{\text{in}}|} \delta\psi \, dx \\ &= - \int_{\Omega} \frac{x}{|\Omega_{\text{in}}|^2} H \, dx \int_{\Omega} \delta\psi \, dx + \int_{\Omega} \frac{x}{|\Omega_{\text{in}}|} \delta\psi \, dx \end{aligned} \quad (29)$$

and

$$\frac{\partial \sigma_{\text{in}}^4}{\partial \phi} = \frac{\partial}{\partial \phi} \left(\int_{\Omega} \frac{(x - \mu_{\text{in}})^2}{|\Omega_{\text{in}}|} H \, dx \right)^2 = 2 \sigma_{\text{in}}^2 \int_{\Omega} \frac{\partial}{\partial \phi} \left[\frac{(x - \mu_{\text{in}})^2}{|\Omega_{\text{in}}|} \right] H + \frac{(x - \mu_{\text{in}})^2}{|\Omega_{\text{in}}|} \delta\psi \, dx. \quad (30)$$

Inserting the various terms into each other, this yields (16).

References

- [1] P. Brodatz. *Textures: A Photographic Album for Artists and Designers*. Dover Publications, 1966.
- [2] R. W. Buccigrossi and E. P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Trans. on Image Proc.*, 8(12):1688–1700, Dec. 1999.

- [3] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *Int. J. of Comp. Vision*, 22(1):61–79, 1997.
- [4] T. Chan and L. Vese. Active contours without edges. *IEEE Trans. on Image Proc.*, 10(2):266–277, Feb. 2001.
- [5] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Interscience, 1991.
- [6] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Trans. Patt. Anal. Mach. Intell.*, 13(9):891–906, September 1991.
- [7] GMBV. First international workshop on generative-model-based vision (GMBV 2002), in conjunction with ECCV 2002. <http://www.diku.dk/publikationer/tekniske.rapper/2002/02-01/>, June, 2 2002. Copenhagen, Denmark.
- [8] D. J. Heeger and J. R. Bergen. Pyramid-based texture analysis/synthesis. In *SIGGRAPH*, pages 229–238, 1995.
- [9] J. Huang and D. Mumford. Statistics of natural images and models. In *Proc. of the ICCV (1)*, pages 541–547, 1999.
- [10] S. Jehan-Besson, M. Barlaud, and G. Aubert. DREAM²S: Deformable regions driven by an Eulerian accurate minimization method for image and video segmentations. *Int. J. of Comp. Vision*, to appear.
- [11] R. L. Joshi, V. J. Crump, and T. R. Fischer. Image subband coding using arithmetic coded trellis coded quantization. *IEEE Trans. on Circuits and Systems for Video Technology*, 5(6):515–523, Dec. 1995.
- [12] S. Kichenassamy, A. Kumar, P. J. Olver, A. Tannenbaum, and A. J. Yezzi. Gradient flows and geometric active contour models. In *Proc. of the ICCV*, pages 810–815, 1995.
- [13] S. M. LoPresto and K. Ramchandran. Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework. In *Proc. of the Data Compr. Conf. (DCC)*, pages 221–230, 1997.
- [14] S. G. Mallat. A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Patt. Anal. Mach. Intell.*, 11(7):674–693, 1988.
- [15] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.
- [16] D. Mumford. Bayesian rationale for the variational formulation. In B. ter Haar Romeny, editor, *Geometry-Driven Diffusion in Computer Vision*, pages 135–146. Kluwer Acad. Publ., Dordrecht, 1994.

- [17] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.
- [18] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *Int. J. of Comp. Vision*, 46(3):223–247, 2002.
- [19] R. Picard, C. Graczyk, S. Mann, J. Wachman, L. Picard, and L. Campbell. VisTex vision texture database. MIT Media Lab.
- [20] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. of Comp. Vision*, 40(1):49–71, 2000.
- [21] J. Puzicha, T. Hofmann, and J. Buhmann. Histogram clustering for unsupervised segmentation and image retrieval. *Pattern Recognition Letters*, 20:899–909, 1999.
- [22] J. Puzicha, Y. Rubner, C. Tomasi, and J. M. Buhmann. Empirical evaluation of dissimilarity measures for color and texture. In *Proc. of the ICCV (2)*, pages 1165–1172, 1999.
- [23] R. C. Reininger and J. D. Gibson. Distributions of the two-dimensional DCT coefficients for images. *IEEE Trans. on Communications*, 31(6):835–839, June 1983.
- [24] J. Simon. Differentiation with respect to the domain in boundary value problems. *Numerical Functional Analysis and Optimization*, 2(7):649–687, 1980.
- [25] E. P. Simoncelli. Statistical models for images: Compression, restoration and synthesis. In *31st Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, 1997. IEEE Sig. Proc. Soc.
- [26] E. P. Simoncelli and E. H. Adelson. Noise removal via Bayesian wavelet coring. In *Third Int'l Conf on Image Proc.*, pages 379–382, Lausanne, 1996. IEEE Sig. Proc. Soc.
- [27] E. P. Simoncelli and W. T. Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *Second Int'l Conf. on Image Proc.*, Washington, DC, 1995.
- [28] J. Sokolowski and J.-P. Zolesio. *Introduction to Shape Optimization*. Springer, New York, 1991.
- [29] A. Srivastava, X. Liu, and U. Grenander. Universal analytical forms for modeling image probabilities. *IEEE Trans. Patt. Anal. Mach. Intell.*, 24(9):1200–1214, 2002.
- [30] J. H. van Hateren and A. van der Schaaf. Independent component filtering of natural images compared with simple cells in primary visual cortex. In *Proc. of the Royal Soc. London*, pages 359–366, 1998.
- [31] Y. Wu, S. Zhu, and X. Liu. Equivalence of Julesz ensembles and FRAME models. *Int. J. of Comp. Vision*, 38(3):247–265, 2000.

- [32] A. Zhu, S.C. and Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Trans. Patt. Anal. Mach. Intell.*, 18(9):884–900, 1996.
- [33] S. Zhu and D. Mumford. Prior learning and Gibbs reaction–diffusion. *IEEE Trans. Patt. Anal. Mach. Intell.*, 19(11):1236–1250, 1997.
- [34] S. Zhu, Y. Wu, and D. Mumford. Filters, random fields and maximum entropy (FRAME). *Int. J. of Comp. Vision*, 27(2):1–20, 1998.